DOCUMENT RESUME

ED 271 514 TM 860 432

AUTHOR Hildebrand, Myrene; Hoover, H. D.
TITLE A Comparative Study of the Reliability and Validity of the "Degrees of Reading Power" and the "Iowa Tests of Basic Skills."
PUB DATE 17 Apr 86
NOTE 19p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (San Francisco, CA, April 17-20, 1986).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Basic Skills; Correlation; Grade 5; Grade 6; Intermediate Grades; *Minimum Competency Testing; Predictive Validity; Reading Comprehension; *Reading Tests; Scores; *Test Reliability; *Test Validity; Vocabulary Skills
IDENTIFIERS *Degrees of Reading Power; *Iowa Tests of Basic Skills

ABSTRACT

This study compared the reliability and validity of two different measures of reading ability, the Degrees of Reading Power (DRP) and the Iowa Tests of Basic Skills (ITBS) Reading test and the ITBS Vocabulary test. The data consisted of scores of 377 grade 5 and grade 6 students on these tests, along with their assigned reading levels in the Macmillan reading series, and teacher-assigned grades from two grading periods. Intercorrelations, including disattenuated values, were the primary method of data analysis. The differences between the ITBS and DRP on the reliability and validity evidence presented were minor and appeared to be of limited practical significance. (Author)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED271514

A comparative study of the reliability and validity of the Degrees of Reading Power and the Iowa Tests of Basic Skills

Presenter: Myrene Hildebrand

Cedar Rapids Community Schools

346 2nd Ave. S.W.

Cedar Rapids, IA 52404

H. D. Hoover

The University of Iowa

316 Lindquist Center

Iowa City, IA 52242

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

D. H. Hoover

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Paper presented at the annual meeting of the

National Council on Measurement in Education

San Francisco - April 17, 1986

TM 860 432

A comparative study of the reliability and validity of the Degrees of Reading Power and the Iowa Tests of Basic Skills

Myrene Hildebrand, Cedar Rapids Community Schools

H. D. Hoover, The University of Iowa

Abstract

This study compared the reliability and validity of two different measures of reading ability, the Degrees of Reading Power (DRP) and the Iowa Tests of Basic Skills (ITBS) Reading test and the ITBS Vocabulary test. The data consisted of scores of 377 grade 5 and grade 6 students on these tests, along with their assigned reading levels in the Macmillan reading series, and teacher-assigned grades from two grading periods. Intercorrelations, including disattenuated values, were the primary method of data analysis. The differences between the ITBS and DRP on the reliability and validity evidence presented were minor and appeared to be of limited practical significance.

Rationale

The primary objective of the study was to examine the reliability and validity of two reading assessment instruments and a measure of vocabulary. The Degrees of Reading Power (DRP) (The College Board, 1983), and the Reading and Vocabulary tests of the Iowa Tests of Basic Skills (ITBS) (Hieronymus, Lindquist, & Hoover, 1978, 1985) were the measures of interest.

The DRP is a recent entry in the field of testing and has enjoyed wide use in its brief history. Of particular interest in this study is the publisher's claim that the test measures a different construct than conventional reading tests. The DRP is characterized as purely a "process test" (Rankin, unpublished manuscipt) designed to measure literal comprehension. As it is explained in the Users Guide, the DRP consists of eleven prose passages of approximately 325 words arranged in ascending order of difficulty. Seven of the sentences in each passage contain blank spaces indicating that a word is missing. Students must select from five single-word options the response that most appropriately fits the blank. Each option would make a gramatically correct and semantically plausible sentence. While each option would make sense in its sentence, only the correct option makes sense in its paragraph (Koslin & Ivens, unpublished manuscript). It is the surrounding text which dictates the plausibility of the response.

The use of information to recognize and understand stated factual details and relationships, to infer underlying relationships, and to develop generalizations from a selection are the constructs upon which the ITBS Reading Comprehension test is built. The explanation in the

Teacher's Guide points out that these skills correspond to literal, interpretative, and evaluative meaning used by many reading authorities to classify objectives of reading. Considered in this way, reading can be equated to thinking and problem solving, made possible by the processing of textual information (Rankin, unpublished manuscript). By comparison then, the DRP addresses a much narrower definition of reading than does the ITBS. The Reading Comprehension test (Test R) samples a much broader domain with 54 items in the Level 11 test and 56 items in the Level 12 test.

The DRP has been heralded as an integrated approach which measures reading comprehension holistically (Rosecrans, 1984). In other words, it does not measure "isolated skills" such as vocabulary. The ITBS, in fact, acknowledges the integral importance of vocabulary in reading comprehension by including items related to contextual meaning within Test R. A separate vocabulary assessment (Test V) contains 39 items in Level 11 and 42 items in Level 12 (41 items in Form G). It is described by the authors as a measure of reading, understanding of word meaning, and also concept development as it relates to linguistic function.

Data and Source

The subjects for the study were all students in the Cedar Rapids Community School District in eastern Iowa. This is a K-12 system with a total enrollment of approximately 18,000 students averaging 1300 per grade levels. The minority student population is about 8%. Of the twenty-one elementary schools, ten are identified as Chapter I. The stability index for the district is .89 ($\frac{\text{initial enrollment - number exiting}}{\text{initial enrollment}}$).

Interest in the intercorrelations between the DRP and the two ITBS tests described above emerged because the ITBS is routinely administered every fall to students in grades two through eight as the major component of the standardized testing program. ITBS scores are available for all students in these grades whose instructional program includes such testing (approximately 95%). The DRP was administered as one of several tests in a matrix sampling during the spring of 1985. The sample of students who took the DRP was determined through a random cluster sampling of classrooms stratified to include a proportionate number of classrooms from identified Chapter I schools. The data were gathered for 191 students in their fifth and sixth grade years and 186 students in their sixth and seventh grade years during the 1984-85 and 1985-86 school years. The students comprise approximately 18% of the total two-grade population in the school district. They represent a cross-section of the ability range of the students-at-large, as verified through examination of Verbal standard age scores which were obtained from the fall testing of each of the students in their fifth grade year. The Cognitive Abilities Test (CogAT) (Thorndike & Hagen, 1971) was used as the ability measure and has been a regularly scheduled component of the district testing program.

The scores from ITBS Vocabulary and Reading Comprehension tests from September 1984 and September 1985 and DRP scores from May 1985 were collected. Correlations between the DRP and ITBS, along with corresponding disattenuated values, were used to establish the concurrent validity between the DRP and each of the ITBS tests of interest. The relative predictive validity of the DRP and ITBS was examined through a

comparison of the intercorrelations between each of these same measures and the reading level to which each of the students was assigned in the Macmillan (1980) reading series in May 1985. In the case of the fifth grade sample, the Macmillan reading level of students as sixth graders in November 1985 was also compared. Comparison was also made with teacher-assigned grades in May 1985 and November 1985.

Results

Means and standard deviations for all variables, along with K-R 20 reliability coefficients for the ITBS and DRP, are given in Table 1. The ITBS grade-equivalent (GE) and CogAT standard age score (SAS) means are very close to the overall district means. The reliability coefficients given in the last column of the table indicate somewhat higher reliabilities for the DRP. This result was not unexpected, given that the DRP contains more items, and that the items on the test are homogeneous (Bruning, 1985). In order to more equitably compare the reliabilities of the tests, the effect of additional test length on reliability can be estimated by application of the Spearman-Brown prophecy formula. The reliability values in parentheses in Table 1 give the estimated values for the ITBS Vocabulary and Reading tests if they each contained the same number of items as the DRP. As can be seen, if these tests were all the same length there would be little to recommend one over the other on the basis of differences in internal consistency reliability.

Since the DRP is an untimed test, it is not possible to directly compare the reliabilities of the tests for a given unit of testing time. However, in practice nearly all students finish the DRP in a 45-50

minute class period. Given the time limits of 15 minutes for the ITBS Vocabulary test and 42 minutes for the ITBS Reading Comprehension test, it would not seem unreasonable to conclude that per given unit of time the ITBS Vocabulary test would be the most reliable, followed by the DRP, and then ITBS Test R. Such a rank ordering would be consistent with the degree of item homogeniety of the three tests.

[Insert Table 1 about here]

Intercorrelations among the ITBS tests and DRP are given in Table 2. As would be expected, all of these correlations are substantial. In addition to the obtained correlations, disattenuated values are reported in this table. These represent estimates of the correlations to be expected between the tests if each were perfectly reliable (K-R 20 = 1.0). Since these are all somewhat smaller than one, they indicate that each of the tests is assessing some unique aspect(s) of reading comprehension.

The disattenuated values between the Vocabulary and Reading tests of the same form of ITBS, that is, those administered during the same testing session for both Form 8 and Form G are slightly higher than those between the DRP and either of the ITBS tests. These higher values would seem to indicate greater similarity in underlying skills being measured. However, the ITBS tests were administered on the same day on the two test occasions, while the DRP administration was separated in time from each of the fall testings--by eight and four months. This time differential is nearly sure to lower the ITBS-DRP correlations considerably.

For the fifth-sixth grade group administered ITBS in September of 1984 and DRP in May of 1985, the obtained DRP-ITBS Vocabulary correlation of .766 was slightly higher than the corresponding DRP-ITBS Reading

Comprehension correlation of .732. (Disattenuated values were .855 and .804 respectively.) For this same group, the obtained correlation between the September 1985 ITBS Vocabulary scores and DRP of .736 was slightly lower than the ITBS Reading Comprehension-DRP correlation of .763. However, the disattenuated values for this comparison were nearly identical, .832 and .839. For the sixth-seventh grade sample, the highest correlation with DRP was with ITBS Reading Comprehension--.818 for the sixth grade year and .806 for the seventh grade year. The DRP correlations with Vocabulary for this group were .763 and .744 for ITBS Forms 8 and G administered in September 1984 and 1985 respectively. These findings differ from the results of a study in which the California Achievement Test (CAT-70) was used as the achievement measure. In that study the DRP correlation with the Reading Comprehension Test (.80) was considerably greater than with DRP and the CAT Vocabulary Test (.56) (Bruning, 1985).

[Insert Table 2 and 3 about here]

Correlations between the Macmillan reading levels to which both groups of students were assigned in May of 1985, and also in the case of the fifth-sixth grade students the assignment of November 1985, and the ITBS national grade-equivalents, and between reading levels and DRP mid-instructional levels, are given in Table 3. Again, these correlations are very similar in magnitude with the ITBS Vocabulary slightly higher in both May and November for the fifth-sixth grade group and for the May assignment of the sixth-seventh grade sample. If the ITBS Vocabulary and Reading scores had been obtained at the same time as the May Macmillan reading levels, as the DRP scores were, those

correlations would probably have been somewhat larger. It would appear that the best predictor of reading series level in this district would be the ITBS Vocabulary test. None of the differences would seem to have much practical significance. In contrast, the DRP is the best predictor of teacher-assigned grades in reading in the spring for the fifth-sixth group and in both spring and fall for the sixth-seventh grade students. However, if differences in reliability of the ITBS and DRP, along with differences between time of testing and grade assignment, are taken into account it would again appear that the obtained differences are of little, if any, practical significance. The low correlations across all tests suggest a lack of reliability in grade assignment, especially in the fall. The comparatively low correlation of all measures with grade reports in the current investigation was found in a study of beginning college freshmen which compared the DRP, ACT Composite scores, and the Nelson-Denny Reading Test (NDRT) with high school grade point average (Wood, Nemeth, & Brooks, 1985). In that study the correlations of the DRP and NDRT with high school GPA were almost negligible.

Discussion

The DRP is designed to be an effectiveness measure of how well the test taker comprehends prose as it is being read. It is confined to the domain of literal comprehension. Because of the identical format of items on the test, the length of the test, and the deliberate delimitations of the domain to information processing, it is not surprising that the test exhibits quite high interval consistency reliabitity.

"Conventional" reading tests such as the ITBS Reading Comprehension test differ from the DRP in the nature of the assessment tasks

necessitated by the broader construct of reading comprehension which includes inferential and evaluative domains. Because these skills subsume literal comprehension, it can be expected that a high correlation would exist between a strict measure of literal comprehension and one which samples from this domain and the others as well. This prediction is well-established by the current study. This strong evidence of convergent validity is reinforced by several other studies. The College Board found a correlation of .80 between the CAT-70 and the DRP, (Bruning, 1985); Wood, Nemeth and Brooks (1985) obtained a .67 correlation between the DRP and the Nelson-Denny Reading Test (NDRT); Weiner and Kippel (1984) found correlations of .77 - .86 between the DRP and the Reading Comprehension test of Forms C and D of the 1979 CAT.

The College Board study found a relatively low correlation between the DRP and scores on the CAT Vocabulary subtest. This would allow for the suggestion that knowledge of vocabulary as such and separate from extended context is not tested by the DRP and, therefore is not be to be considered as an indicant of the DRP construct of literal comprehension. If such correlational evidence would support a claim for discriminant validity, the high correlations between the DRP and ITBS Vocabulary test in this study would suggest the opposite.

The unaccounted for variance between CAT-70 Vocabulary and DRP ($r = .56$) is 69%. This is narrowed to a range from 31-45% ($r = .74 - .83$) in the four correlations obtained in the present study with ITBS Test V and DRP. These are very close to the correlations between the DRP and ITBS Test R in which the unaccounted for variances

range from 34-47% (r = .73 - .82). This compares with 36% (r = .80) between CAT-70 Reading and DRP. The disattenuated values calculated for each of the ITBS-DRP correlations would suggest even less unexplained variance. The important role of knowledge of vocabulary can certainly not be dismissed in the consideration of literal comprehension. The present study clearly brings discriminant validity into question.

The high internal consistency reliabilities of both the DRP and ITBS, and their suggested concurrent validity, support the probability that ranking students within groups would be predictable from one test to the other. This parallels the conclusion of a study of DRP-CAT Reading scores--that results from the two measures would provide educators with functional--equivalent information (Weiner & Kippel, 1984). On the basis of the reliability and validity evidence presented in the preceeding results section, it would seem there is little reason to prefer the ITBS Vocabulary or Reading Comprehension tests over the DRP or vice versa. Concerns beyond the scope of this investigation suggest that attention to other variables might very well provide reason for preference. These include the following: Would different predictions regarding reading performance result if the norms provided with the tests are used to make instructional decisions, as is suggested by the study by Wood, Nemeth, and Brooks (1985)? Is it possible to generalize reading performance on nonfiction to other genre, as questioned by Hanna (1985)? Does achievement in a narrowly defined domain provide an adequate model for comprehensive instruction in reading? Is use of student/teacher time warranted for multiple assessments, one of which provides information on a subset of another? Such practical considerations as these should guide responsible decision-making.

Bibliography

Bruning, R. (1985). Review of Degrees of Reading Power. The Ninth Mental Measurements Yearbook. James Mitchell Jr., ed. Buros Institute of Mental Measurement, University of Nebraska, Lincoln.

Hanna, G. S. (1985). Review of Degrees of Reading Power. The Ninth Mental Measurements Yearbook. James Mitchell Jr., ed. Buros Institute of Mental Measurement, University of Nebraska, Lincoln.

Hieronymus, A. N., Lindquist, E. F., & Hoover, H. D. (1979). Iowa Tests of Basic Skills teacher's guide. Riverside Publishing Company, Chicago.

Hieronymus, A. N., Hoover, H. D., & Lindquist, E. F. (1986). Iowa Tests of Basic Skills preliminary teacher's guide. Riverside Publishing Company, Chicago.

Koslin, B. L., & Ivens, S. H. (unpublished manuscript). DRP validity revisited: A rejoinder.

Rankin, E. F. (unpublished manuscript). The validity of the Degrees of Reading Power test: A response to Carver's critique.

Rosecrans, D. (1984). One college's use of the Degrees of Reading Power test in a developmental reading program. College Entrance Examination Board.

Users Guide PA Series. (1983). Degrees of Reading Power. The College Board.

Weiner, M., & Kippel, G. (1984). The relationship of the California Achievement Test to the Degrees of Reading Power test. Educational and Psychological Measurement, 44, 497-500.

Wood, P., Nemeth, J. S., & Brooks, C. C. (1985). Criterion related validity of the Degrees of Reading Power test (Form CP-14). Educational and Psychological Measurement, 45(4), 965-969.

Table 1

ITBS, DRP, CogAT, and Reading Grades Summary Statistics
Grade 5-6 Sample, N = 191

Test	Administration Date	# of Items	Score	Mean	SD	Reliability (K-R 20)
ITBS-Vocabulary (Form 8-Level 11)	Sept. '84	39	RS	25.4	7.3	.869
			GE	60.1	13.0	(.929)**
ITBS-Reading (Form 8-Level 11)	Sept. '84	54	RS	31.6	9.7	.898
			GE	57.7	14.6	(.926)
CogAT-Verbal	Sept. '84	100	SAS	108.1	12.8	
DRP (Form PA6)	May '85	77	RS	49.3	12.1	.924
			MIL*	55.4	11.2	
Reading Grades	June '85		***	8.4	2.3	
ITBS-Vocabulary (Form G-Level 12)	Sept. '85	41	RS	23.5	7.0	.847
			GE	65.6	12.5	(.912)
ITBS-Reading (Form G-Level 12)	Sept. '85	56	RS	32.0	9.9	.896
			GE	65.4	15.1	(.922)
Reading Grades	Nov. '85		***	8.2	2.2	

* MIL = Mid Instructional Level, DRP Readability Scale

** Estimates of reliability of ITBS tests if each had the same number of items as the DRP (77 items)

*** Grade scale, F = 1, D- = 2, ..., C = 6, ..., A = 12, A+ = 13.

Table 1 (continued)

ITBS, DRP, CogAT, and Reading Grades Summary Statistics
Grade 6-7 Sample, N = 186

Test	Administration Date	# of Items	Score	Mean	SD	Reliability (K-R 20)
ITBS-Vocabulary	Sept. '84	42	RS	26.9	8.8	.904
(Form 8-Level 12)			GE	68.0	14.5	(.961)**
ITBS-Reading	Sept. '84	56	RS	33.3	11.5	.921
(Form 8-Level 12)			GE	69.9	17.1	(.941)
CogAT-Verbal	Sept. '83	100	SAS	106.5	13.7	
DRP	May '85	77	RS	59.9	13.5	.947
(Form PB6)			MIL*	61.5	14.0	
Reading Grades	June '85		***	9.0	2.3	
ITBS-Vocabulary	Sept. '85	41	RS	23.6	7.5	.859
(Form G-Level 13)			GE	75.2	17.9	(.919)
ITBS-Reading	Sept. '85	57	RS	30.7	11.2	.918
(Form G-Level 13)			GE	74.5	20.9	(.938)
Reading Grades	Nov. '85		***	7.5	2.8	

* MIL = Mid Instructional Level, DRP Readability Scale

** Estimates of reliability of ITBS tests if each had the same number of items as the DRP (77 items)

*** Grade scale, F = 1, D- = 2, ..., C = 6, ..., A = 12, A+ = 13.

Table 2

Intercorrelations among ITBS Grade Equivalent and DRP Mid Instructional Level Scores Grade 5-6 Sample

	ITBS Vocabulary 5-Fall	ITBS Reading 5-Fall	DRP 5-Spring	ITBS Vocabulary 6-Fall	ITBS Reading 6-Fall
ITBS-Vocabulary 5-Fall	----	.792	.766	.807	.784
ITBS-Reading 5-Fall	.897	----	.732	.684	.747
DRP 5-Spring	.855	.804	----	.736	.763
ITBS-Vocabulary 6-Fall	.941	.784	.832	----	.752
ITBS-Reading 6-Fall	.888	.833	.839	.863	----

(Obtained correlations are given above the diagonal, disattenuated values below the diagonal.)

Table 2 (continued)

Intercorrelations among ITBS Grade Equivalent and DRP Mid Instructional Level Scores Grade 6-7 Sample

	ITBS Vocabulary 6-Fall	ITBS Reading 6-Fall	DRP 6-Spring	ITBS Vocabulary 7-Fall	ITBS Reading 7-Fall
ITBS-Vocabulary 6-Fall	----	.812	.763	.844	.811
ITBS-Reading 6-Fall	.890	----	.818	.788	.853
DRP 6-Spring	.825	.869	----	.744	.806
ITBS-Vocabulary 7-Fall	.958	.875	.825	----	.779
ITBS-Reading 7-Fall	.890	.928	.864	.877	----

(Obtained correlations are given above the diagonal, disattenuated values below the diagonal.)

Table 3

Intercorrelations Between Test Scores and Grades and
Between Test Scores and Macmillan Reading Levels
Grade 5-6 Sample

	Reading Grades		Macmillan-Reading Levels	
	5-Spring	6-Fall	5-Spring	6-Fall
ITBS-Vocabulary	.529	.412	.602	.604
5-Fall	(.567)*	(.441)	(.646)	(.648)
ITBS-Reading	.545	.380	.564	.548
5-Fall	(.575)	(.401)	(.595)	(.578)
DRP	.615	.437	.552	.550
5-Spring	(.640)	(.455)	(.574)	(.572)
ITBS-Vocabulary	.466	.360	.606	.592
6-Fall	(.506)	(.391)	(.658)	(.643)
ITBS-Reading	.605	.511	.555	.519
6-Fall	(.639)	(.540)	(.586)	(.548)

* Values in parentheses are estimates of intercorrelations if both ITBS and DRP were perfectly reliable tests.

Table 3 (continued)

Intercorrelations Between Test Scores and Grades and
Between Test Scores and Macmillan Reading Levels
Grade 6-7 Sample

	Reading Grades		Macmillan Reading Levels
	6-Spring	7-Fall	6-Spring
ITBS-Vocabulary	.547	.358	.582
6-Fall	(.575)	(.377)	(.612)
ITBS-Reading	.615	.409	.558
6-Fall	(.640)	(.426)	(.581)
DRP	.632	.527	.600
6-Spring	(.649)	(.542)	(.617)
ITBS-Vocabulary	.555	.395	.620
7-Fall	(.599)	(.426)	(.669)
ITBS-Reading	.599	.419	.592
7-Fall	(.625)	(.437)	(.618)